Discovery of Potential 1,3,5-Triazine Compounds Against Strains of Plasmodium falciparum Using Supervised Machine Learning Models

Supriya Sahu , Surajit Kumar Ghosh , Jun Moni Kalita , Murali C Ginjupalli , Kranthi Raj K

PII: S0928-0987(19)30481-6
DOI: https://doi.org/10.1016/j.ejps.2019.105208
Reference: PHASCI 105208

To appear in: *European Journal of Pharmaceutical Sciences*

Received date: 10 October 2019
Revised date: 29 November 2019
Accepted date: 22 December 2019

Please cite this article as: Supriya Sahu , Surajit Kumar Ghosh , Jun Moni Kalita , Murali C Ginjupalli , Kranthi Raj K , Discovery of Potential 1,3,5-Triazine Compounds Against Strains of Plasmodium falciparum Using Supervised Machine Learning Models, *European Journal of Pharmaceutical Sciences* (2019), doi: https://doi.org/10.1016/j.ejps.2019.105208

Discovery of Potential 1,3,5-Triazine Compounds Against Strains of *Plasmodium falciparum* Using Supervised Machine Learning Models

Supriya Sahu*[a] Surajit Kumar Ghosh[a] Jun Moni Kalita[a] Murali C Ginjupalli[b] and Kranthi Raj K[c]

[a] *Department of Pharmaceutical Sciences, Dibrugarh University, Dibrugarh 786004 Assam, India.*
[b] *CaroCure Discovery Solutions Pvt. Ltd., 2897 Churchhill Lane Saginaw MI 48603, USA.*
[c] *CaroCure Discovery Solutions Pvt. Ltd., IKP Knowledge Park, Genome Valley, Shamirpet, Hyderabad-500 101 Telangana India.*

*Corresponding author: supsjrt@gmail.com

Abstract

The Malaria burden was an escalating global encumbrance and need to be addressed with critical care. Anti-malarial drug discovery was integrated with supervised machine learning (ML) models to identify potent thiazolyl-traizine derivatives. This assimilated approach of Direct Kernel-based Partial Least Squares regression (DKPLS) with molprint 2D fingerprints in Quantitative Structure Activity Relationship models was utilized to map the knowledge of known actives and to design novel molecules. This QSAR study had revealed the structural features required for better antimalarial activity. Two of the molecules which were designed based on the results of this QSAR study, had shown good percentage of parasitemia against both chloroquine sensitive (3D7) and chloroquine resistant (Dd2) strains of *Plasmodium falciparum* respectively. The IC_{50} of 201D and 204D was 3.02 and 2.17 µM against chloroquine resistant Dd2 strain of *Plasmodium falciparum*. This result had proved the efficiency of a multidisciplinary approach of medicinal chemistry and machine learning for the design of novel potent anti-malarial compounds.

Keywords:

Anti-malarials
1, 3, 5-triazine
Plasmodium falciparum
Machine Learning
QSAR

1. Introduction

Based on the observation that Cycloguanil which was a potent antimalarial antifolate inhibits *Lactobacillus casei* moderately [1, 2] the synthesis of 2-amino-4-phenyl thiazolyl-1, 3, 5- triazine derivatives as antimalarial was started. Some of the thiazolyl-triazine derivatives synthesized by our group for the first time showed encouraging inhibition of *Lactobacillus casei* [3-6]. Testing the compounds against *Lactobacillus casei* clarifies the observation found between antibacterial and antimalarial effect of Cycloguanil. These observations inspired us to design different thiazolyl-traizine derivatives and to explore their antimalarial activities. Till now our group has synthesized hundreds of such derivatives out of which, the thiazolyl-triazine core having secondary amine substitution on the triazine ring and chlorine or nitro substitution on the second, third and/or fourth positions of the phenyl thiazole ring (Table 1) exhibited better activity compared to the others [7, 8]. This finding is taken as the backbone of the current study which is further analysed using machine learning (ML) models.

ML is the exercise of expending systems to analyze statistics, acquire knowledge from it and mark a determination. It is also a forecast of the upcoming result of original data. Here the machine is taught by means of statistics and algorithms that contribute it the capability to learn to accomplish the assignment. Supervised and unsupervised learning are the two types of methods that can be used based on the available data. Supervised learning approaches are used to progress training datasets to calculate forthcoming values and these are all constant variables, while unsupervised procedures are used for investigative drives to improve models that permit assembling of the statistics that are not indicated by the user. So in this study we used supervised learning methods that train a model on known data that can predict impending outputs using known inputs. The presence of in house generated data on antimalarial activity prompted us to study 2D and 3D QSAR models for further refinement in their structure and enhanced antimalarial activity.

Fifty seven thiazolyl triazine derivatives which has been reported by our group previously in various journals [7-10] are been used for this QSAR study. Based on the results of the QSAR studies favourable substitutions on 1, 3, 5-triazine ring was made with piperazine ring. The overlap of the 2D and 3D visualization shows that this group has favourable effect on the activity. These analogues with piperazine on triazine ring was synthesized and tested for the biological activity. 201D and 204D had shown promising antimalarial activity against the Chloroquine resistant strain of *Plasmodium falciparum* (Dd2).

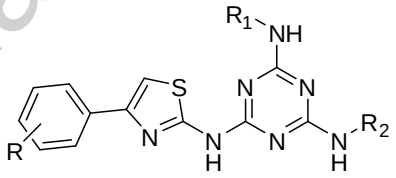
2. Material and methods

2.1 2D & 3D QSAR

Data-driven or Machine learning methods are rapidly evolving and becoming very important in various stages of drug discovery. The Supervised machine learning models are very significant in generating the QSAR models. In this work we have applied supervised machine learning methods for the 2D and 3D finger-prints of the anti-malarial compounds with respect to their biological activity to generate 2D and 3D QSAR models respectively.

The first and significant step in these data driven method was the selection of the input samples (Training Set). Using supervised method, hierarchical clustering was used to cluster 57 antimalarial compounds according to their chemical space (Table 1). Radial hashed fingerprint was used for the clustering according to their chemical space. Compounds from each cluster with different ranges of activity were selected as training set to cover chemical space and activity range. Thus 47 out of 57 compounds were selected as training set having activity (pIC_{50}) range of 5.0 to 2.82 to build the 2D QSAR model. Remaining 10 compounds with activity (pIC_{50}) range of 4.92 to 2.82 were used as test set to validate the hypothesis. There was a small change in the training and test sets of the 2D and 3D QSAR models. This is because the 2 molecules alignment was not proper for the 3D QSAR and so those were added to test set.

Table 1: List of compounds used in QSAR models



Sl No.	Compound Code	Structure	Activity	Test/Training 2D	Test/Training 3D
1	A5	R= 2, 4-dichloro; R_1, R_2= -CH_3	4.999	Training	Training
2	H16	R= 3-nitro; R_1, R_2= -Phenyl	4.945	Training	Training
3	A3	R= 2, 4-dichloro; R_1, R_2= -n-propyl	4.904	Test	Test
4	H18	R= 3-nitro; R_1, R_2= -piperidinyl	4.874	Training	Training
5	H2	R= 3-nitro; R_1, R_2= -cyclopropyl	4.818	Training	Training
6	A29	R= 2, 4-dichloro; R_1, R_2= -benzyl	4.736	Training	Training
7	66B	R= 4-chloro; R_1=-H, R_2= piperazinyl	4.65	Training	Training
8	A2	R= 2, 4-dichloro; R_1, R_2= -cyclopropyl	4.632	Test	Test
9	A21	R= 2, 4-dichloro; R_1, R_2= -o-toluidininyl	4.586	Training	Training

10	A20	R= 2, 4-dichloro; R_1, R_2= -p-toluidininyl	4.539	Training	Test
11	79B	R= 4-chloro; R_1=aminoethyl, R_2= $-CH_3$	4.511	Training	Training
12	A19	R= 2, 4-dichloro; R_1, R_2= -cyclohexyl	4.509	Training	Training
13	68B	R= 4-chloro; R_1=hydroxyethyl, R_2= -H	4.488	Training	Training
14	H5	R= 3-nitro; R_1, R_2= $-CH_3$	4.487	Training	Test
15	H20	R= 3-nitro; R_1, R_2= -p-toluidininyl	4.483	Training	Training
16	H19	R= 3-nitro; R_1, R_2= -cyclohexyl	4.481	Test	Training
17	H29	R= 3-nitro; R_1, R_2= -benzyl	4.472	Training	Training
18	81B	R= 4-chloro; R_1=$-CH_3$, R_2= piperazinyl	4.45	Test	Training
19	A4	R= 2, 4-dichloro; R_1, R_2= -dimethyl	4.447	Training	Training
20	E16	R= -H; R_1, R_2= -Phenyl	4.421	Training	Training
21	A16	R= 2, 4-dichloro ; R_1, R_2= -Phenyl	4.417	Training	Training
22	A18	R= 2, 4-dichloro; R_1, R_2= -piperidinyl	4.415	Training	Test
23	E3	R= -H; R_1, R_2= -n-propyl	4.41	Training	Training
24	E2	R= -H; R_1, R_2= -cyclopropyl	4.391	Training	Training
25	H4	R= 3-nitro; R_1, R_2= -dimethyl	4.383	Training	Training
26	E5	R= -H; R_1, R_2= $-CH_3$	4.373	Test	Test
27	H3	R= 3-nitro; R_1, R_2= -n-propyl	4.34	Training	Training
28	H21	R= 3-nitro; R_1, R_2= -o-toluidininyl	4.263	Test	Test
29	F5	R= 4-chloro; R_1, R_2= $-CH_3$	4.253	Training	Training
30	67B	R= 4-chloro; R_1= -H, R_2= methylpiperazinyl	4.15	Training	Training
31	78B	R= 4-chloro; R_1= $-CH_3$, R_2= morpholinyl	3.85	Training	Training
32	34d	R=3,4-dichloro; R_1= diethyl,R_2= $-SHC_4H_9$	3.59	Training	Training
33	E18	R= -H; R_1, R_2= -piperidinyl	3.56	Training	Test
34	E20	R= -H; R_1, R_2= -p-toluidininyl	3.56	Training	Training
35	F21	R= 4-chloro; R_1, R_2= -o-toluidininyl	3.47	Training	Training
36	84B	R= 4-chloro;R_1= $-CH_3$,R_2= hydroxyethylpiperazinyl	3.47	Training	Training
37	F2	R= 4-chloro; R_1, R_2= -cyclopropyl	3.44	Training	Training
38	F18	R= 4-chloro; R_1, R_2= -piperidinyl	3.44	Training	Training
39	23c	R= 4-chloro; R_1= diethyl, R_2= furan-2-methyl	3.44	Training	Training
40	82B	R= 4-chloro; R_1= $-CH_3$, R_2= methylpiperazinyl	3.31	Training	Training
41	E21	R= -H; R_1, R_2= -o-toluidininyl	3.2	Training	Test
42	F3	R= 4-chloro; R_1, R_2= -n-propyl	3.2	Test	Training
43	28d	R=3,4-dichloro; R_1= diethyl, R_2= -Cl	3.2	Training	Training
44	21c	R= 4-chloro; R_1= diethyl, R_2= -n-butyl	3.13	Test	Test
45	12b	R= 3, 4-dichloro; R_1= -H, R_2= -n-butyl	3.11	Training	Training
46	F4	R= 4-chloro; R_1, R_2= -dimethyl	3.1	Training	Training
47	F19	R= 4-chloro; R_1, R_2= -cyclohexyl	3.1	Training	Test
48	72B	R= 4-chloro; R_1= -H, R_2= -H	2.99	Test	Training

49	4a	R= 4-chloro; R_1= -H, R_2= -phenyl	2.98	Training	Training
50	F20	R= 4-chloro; R_1, R_2= -p-toluidininyl	2.98	Training	Training
51	29d	R=3,4-dichloro; R_1= diethyl, R_2= isopropyl	2.94	Training	Training
52	35d	R=3,4-dichloro; R_1= diethyl, R_2= -SHC_6H_5	2.92	Training	Training
53	69B	R= 4-chloro; R_1= -H, R_2= hydroxyethylpiperazinyl	2.82	Training	Training
54	E4	R= -H; R_1, R_2= -dimethyl	2.82	Training	Test
55	E19	R= -H; R_1, R_2= -cyclohexyl	2.82	Training	Training
56	E29	R= -H; R_1, R_2= -benzyl	2.82	Training	Training
57	F29	R= 4-chloro; R_1, R_2= -benzyl	2.82	Test	Training

Numerical representation of the structural features and atom typing schemes of input compounds required for the quantitative prediction using QSAR techniques. The hashed fingerprints, Molprint2D were used to generate the binary fingerprint for all the compounds used in this study. Molprint2D considers each heavy atom in a structure and is characterized by an environment that consists of all other heavy atoms within a distance of two bonds [11]. We have used supervised learning method that couples Direct Kernel-based Partial Least Squares regression (DKPLS) [12] with molprint2D fingerprints in Canvas software [13] to obtain predictive and interpretable 2D QSAR models. Comparison with other eigenvector methods, DKPLS is highly efficient and its low rank approximation tends to yield more robust regression models that can be visualized in terms of favourable and unfavourable structural characteristics on the molecule. It produces the low rank approximation which is suited well to regressions and in turn gives accurate results. Orthogonal factorization is used by DKPLS regression to build low ranked estimates of kernel matrix K. These approximation and estimates were used to evaluate regression in this work. Total emphasis was on how this technique is adjusted to use through fingerprints generated using molprint along with QSAR visualization and cunning of uncertainties in the estimates of the models. Gaussian kernel matrix with auto scaled variable pairs denoted by column vectors X_i & X_j is defined as follows.

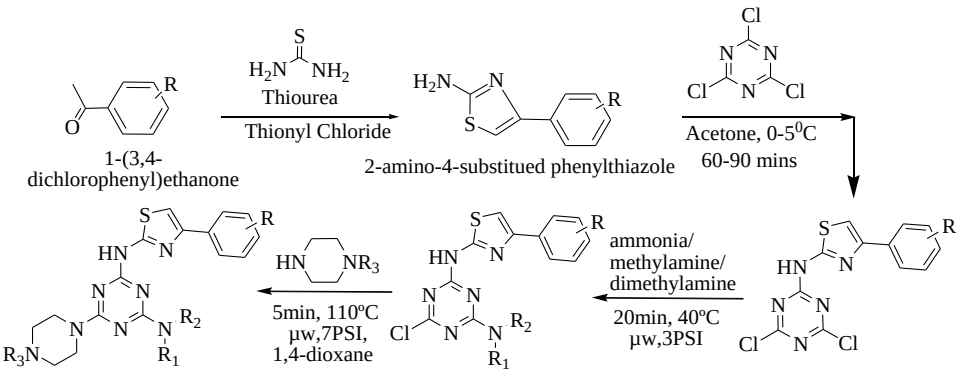
$$K_{ij} = \exp(-|| X_i - X_j ||2 / 2\sigma 2$$

Atom based 3D QSAR models were developed based on our previous work on antimalarials [14].

2.2 Chemistry

The QSAR study performed in this work was considered as the basis for the design of new potent thiazolyl-triazine derivatives as antimalarial compounds. These compounds were synthesized using scheme 1.

Scheme 1: Synthesis of thiazolyl-traizine derivatives



Compound code	R	R_1	R_2	R_3
201D	3, 4-Dichloro	H	CH_3	H
204D	3, 4-Dichloro	H	CH_3	CH_2CH_2OH
7A	H	H	H	CH_3
144C	4-CH_3	H	CH_3	CH_2CH_2OH
171C	4-CH_3	CH_3	CH_3	H

2.3 Antimalarial activity evaluation

The chloroquine sensitive 3D7 and chloroquine resistant Dd2 strain of *P. falciparum* was routinely maintained in medium, rose well park memorial institute-1640 (RPMI-1640) supplemented with 25 mmol 4-(2-hydroxyethyl)-1-piperazin-ethane sulphonic acid (HEPES), 1% D(+)-glucose, 0.23% sodium bicarbonate and 10% heat inactivated human serum. The asynchronous parasites of *P. falciparum* were synchronized after 5% D-sorbitol treatment to obtain only the ring stage parasitized cells. The initial ring stage parasitaemia of 0.8–1.5% at 3% haematocrit in a total volume of 200 µL of medium RPMI-1640 was uniformly maintained for carrying out the assay.

The *in-vitro* antimalarial assay was carried out with modified micro-assay of Reickmann and co-workers [1] in 96 well-microtitre plates. A stock solution of 5 mg/mL of each of the test samples was prepared in DMSO and subsequent dilutions were made with culture medium. The test compounds in 20 µL volume at 50 µg/mL concentration in duplicate well were incubated with parasitized cell preparation at 37°C and 5% CO_2 level in a carbon dioxide incubator. After 36–40 h of incubation, the blood smears were prepared from each well and stained with 3% Giemsa. The level of parasitemia in terms of % dead rings along with schizonts was determined by counting a total of 100 asexual parasites microscopically. Chloroquine was used as reference standard drug. Table 4 shows the *in-vitro* antimalarial activity of the synthesized molecules.

3. Results and Discussion

3.1 QSAR modelling

3.1.1 2D-QSAR modelling

The QSAR model was generated using 47 training set compounds using KPLS method with a maximum of four latent factors based on standard deviation of regression (SD). The model was validated using 10 test set compound for evaluating the predictive power of the model. The statistics of the generated model was shown in the table 2. As listed in the table, at four factor KPLS model showed a good internal correlation (R^2) of 0.79 between the actual activity and predicted activity (figure 1) for the training set. The model is having high predictive power for the external dataset having a correlation (Q^2) of 0.78.

Table 2: 2D QSAR statistics for the 4 latent variables generated using KPLS method

Factors	1	2	3	4
SD	0.59	0.43	0.36	0.34
R^2	0.33	0.65	0.75	0.79
Q^2	0.28	0.64	0.72	0.78
RMSE	0.59	0.42	0.36	0.33

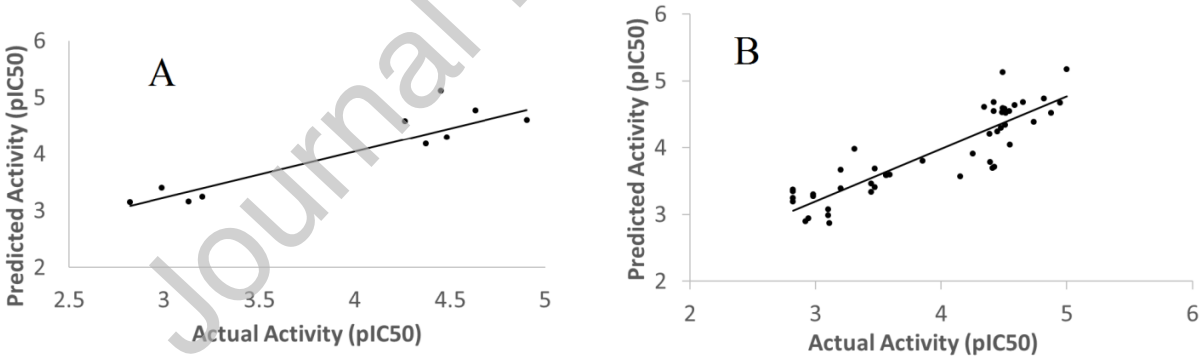


Figure 1: A) Actual and predicted activity of training set compounds obtained in 2D QSAR. B) Actual and predicted activity of test set compounds obtained in 2D QSAR.

3.1.2 2D QSAR Model Visualization

A visual analysis of atomic effects for Canvas KPLS models built from molprint 2D were shown in Figure 2 for compounds with strong (H-bonded interaction) and weak affinities (ionic interaction) for the antimalarial compounds in this study. Favourable atoms for the activity were coloured red, whereas non-favourable atoms were coloured blue. Further the colour intensity reflects the strength of the effect.

In the highest active compound A5, the two amino methyl substitutions on the 1,3,5-triazine and the second Carbon of di-chloro substituted phenyl were mapped with high intense red color indicating that these groups were important for strong binding of antimalarial compounds with the target. Another active compound H18, the anionic oxygen of NO_2 group substituted on meta- position of phenyl ring was showing strong red colour and contributed for higher activity. The nitrogen atom of the two piperidine ring substituted on the 1,3,5-triazine ring was also contributing for the activity. The hydrophobic carbon (blue color) of piperadine ring was not favouring the activity. This trend was also observed in least active compounds F29 and 29C, the hydrophobic atoms on these compounds were showing high intense blue colour and reduced activity.

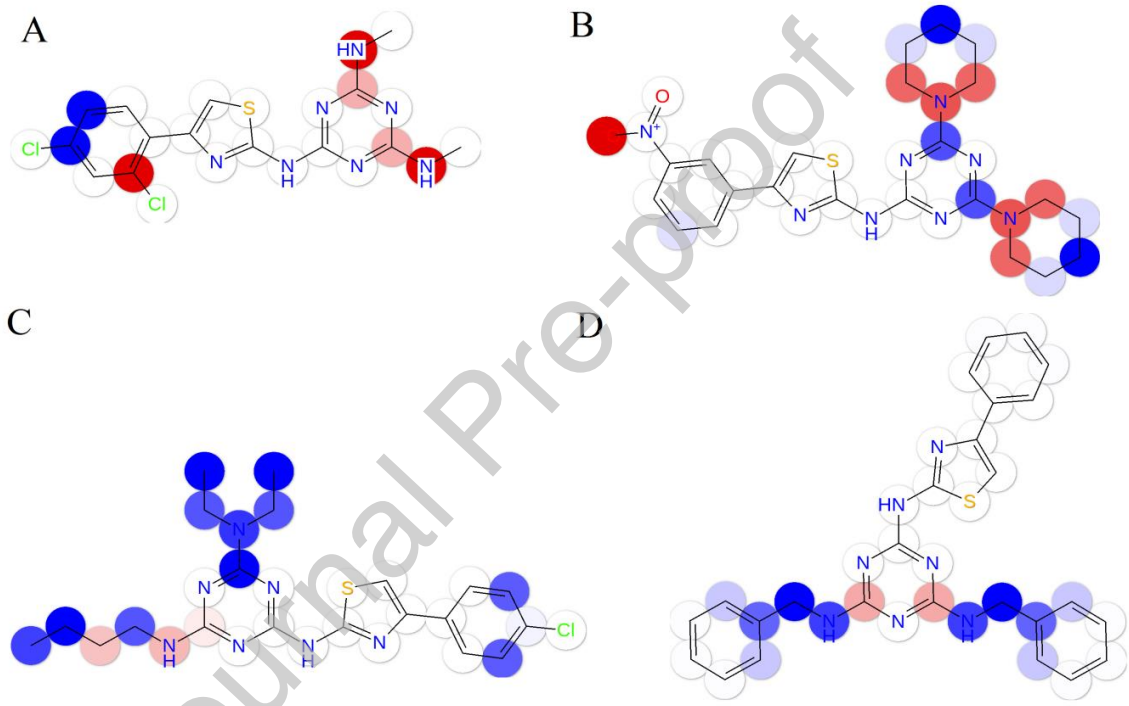


Figure 2: 2D-QSAR visualization of highest active compounds A5 (A), H18 (B) and least active compounds 29d (C), F29 (D). Atoms mapped with red are favourable for activity and the blue are non-favourable for the activity.

3.1.3 3D-QSAR Modelling

Further to explore the 3D features of the substitutions on the phenyl thiazol-2-yl-1,3,5-triazine-2 amine scaffold and to quantify the activity, a supervised learning model using atom based properties a 3D QSAR model was generated. During the generations of atom-based QSAR, all the compounds were aligned according to similarity in their shape with respect to the highest active compound. Based on the occupancy of type of atom of each aligned molecule in 3 dimensional grid space quantitative models were generated. The types of atoms were hydrogen bond donors, hydrophobic/non-polar (H), negative

ionic (N), positive ionic (P) and electron-withdrawing (W) atoms. Each atom occupied in the grid space was converted into bit string. The bit string was a collection of binary valued 3D descriptors and considered as independent variables to generate a quantitative model using partial least squares (PLS). For generating the models, 57 antimalarial compounds synthesized earlier by our group were divided into 45 compounds as training set and 12 compounds as test set using unsupervised learning methods (described elsewhere in the paper). Using 5 latent variables and with leave-one-out (LOO) cross validation method 3D QSAR model was generated and the predictive statistics was illustrated in the table 3 and figure 3. As show in table 3, 5 factorial models were having the highest internal correlation (R^2) of 0.82 and also external correlation of 0.81.

Table 3: Atom Based 3D-QSAR statistics for the 5 latent variables generated by using PLS method

Factors	1	2	3	4	5
SD	0.58	0.44	0.37	0.34	0.33
R^2	0.37	0.65	0.75	0.8	0.82
R^2 CV	0.26	0.45	0.55	0.59	0.6
P	5.52^{e-08}	1.29^{e-10}	6.26^{e-13}	5.49^{e-14}	4.4^{e-14}
RMSE	0.66	0.4	0.31	0.33	0.3
Q^2	0.1	0.66	0.8	0.78	0.81
Pearson-r	0.34	0.85	0.93	0.95	0.95

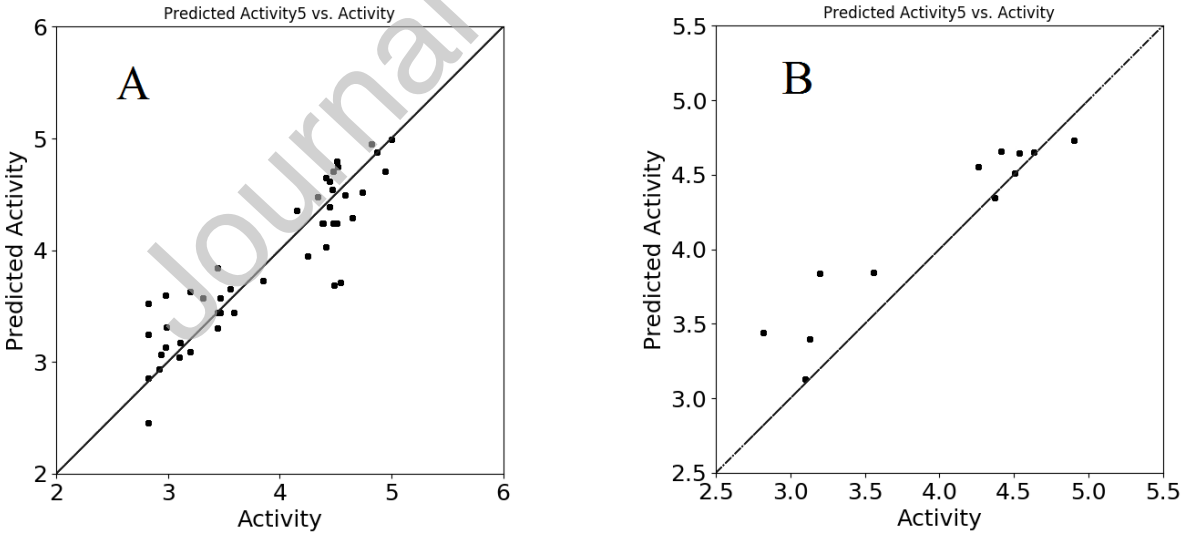


Figure 3: The scatter plot between the actual activity (X-axis) vs predicted activity (Y-axis) obtained by using 5 latent variables of atom based QSAR. A is the training set correlation and B is the test set correlation.

3.1.4 3D-QSAR Visualisation

The 3D-countours generated by the atom based QSAR model showed the significance of the substituents on the biological activity. Figure 4 illustrates the most significant favourable (blue) and unfavourable interactions (red) when the two-factor QSAR model was applied to active ($IC_{50} > 4.2$ µM) and inactive compounds ($IC_{50} < 3.2$ µM).

The large blue region around potential hydrogen bond donors on the substitutions of 1, 3, 5-triazine ring suggest that these features were important for high activity. The compounds 66B, 79B were showing higher activity due to the presence of donor group at this position. The blue contours of electronegative, electro positive and electron withdrawing on the active compounds were overlapping with the NO_2 group of highly active compounds H16, H18 and H2. The small blue colour region near 2 and 4 substitution of 1, 3, 5-triazine ring suggested that smaller hydrophobic groups were tolerable at this region and hence the N-methyl substitutions of the compound A5 and A3 were showing highest activity. A large red region depicted the non-favourable interaction for activity. Most of the inactive compounds were substituted with large hydrophobic groups at these positons and hence they were weak binders. The blue colour path near ortho and meta-position of phenyl ring showed the favourable region of the hydrophobic groups. The presence of the chlorine atom at this position tends to increase the activity. But the non-favourable hydrophobic feature at para position of the phenyl ring decreases the antimalarial activity for the compounds.

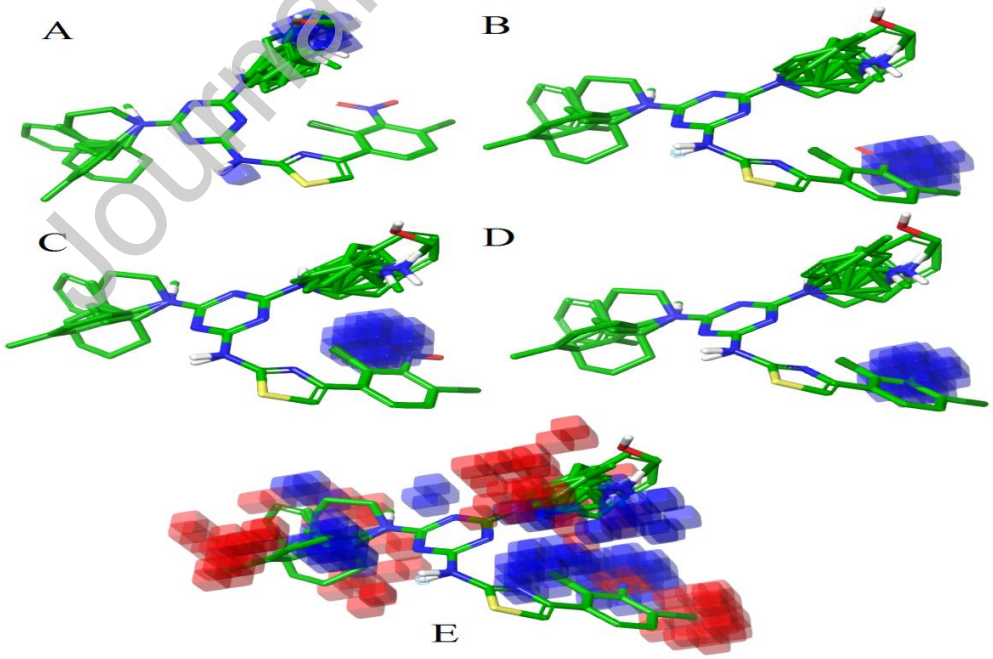


Figure 4: A- Hydrogen Bond Donor, B- electronegative, C- electron withdrawing, D- electro positive, E-, electro positive and hydrophobic maps on active compounds. Blue colour contours indicates the favourable regions and red colour indicates the non-favourable regions.

3.1.5 Visualizations of 2D and 3D QSAR models on newly designed compounds

Based on the results of the 2D and 3D QSAR studies, favourable nitrogen atom was kept on 1, 3, 5-triazine ring and the non-favourable hydrophobic atoms were substituted with favourable hydrogen bond donors like substituted piperazine. We have tried piperazine bioisosteres like 2,6-diazaspiro [3.3] heptane, (1S,4S)-2,5-diazabicyclo [2.2.1] heptane and octahydropyrrolo [3,4-c] pyrrole. Among all the bioisosteres piperazine moiety has good synthetic feasibility, so preceded with piperazine ring. Further as detected from 3D QSAR study, the favourable hydrophobic region at meta position of the phenyl ring was obtained by substituting the hydrogen with chlorine atom to design new analogues. The overlap of the 2D and 3D visualization (Figure 5) showed that this group has favourable effect on the activity. Two analogues with piperazine on triazine ring and chlorine at meta position was synthesized and tested for their *in-vitro* antimalarial activity. Further to validate our hypothesis we synthesized 3 more compounds which were predicting low activity with both the QSAR models and were tested.

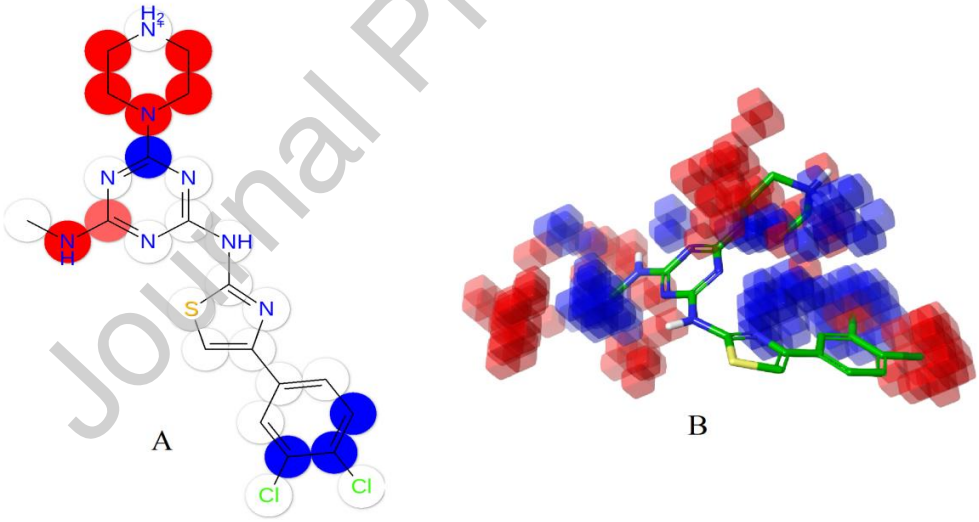


Figure 5: Visualizations of 2D and 3D QSAR models on newly designed compounds; A. 2D visualization map; B are the countor maps of 3D QSAR on the new design.

3.2 Chemistry

N2-(4-(3,4-dichlorophenyl)thiazol-2-yl)-N4-methyl-6-(piperazin-1-yl)-1,3,5-triazine-2,4-diamine (**201D**): Physical state: yellow powder; % yield: 71; m.p.: 280ºC; R_f (silica gel G): 0.66 (ethyl

acetate:hexane, 1:1); Solubility: DMSO, ethanol, methanol, chloroform. UV_{max} (nm): 290.0; FTIR (cm^{-1}) 3369.20 (N-H sec, Str.); 3103.60 (C-H Str.); 1237.21, 1182.08 (CN, Ar.) ^{1}H NMR (MeOD): δ ppm: 2.5 (s, 3H, methyl), 2.9 (d, 4H, piperazine), 3.4 (d, 4H, piperazine), 6.7 (s, 1H, thiazole), 7.5 (d, 2H, phenyl), 7.8 (d, 2H, phenyl). ^{13}CNMR (MeOD): δ, ppm: 40.01, 122.38, 124.73, 126.67, 128.22, 129.40, 149.01, 152.21, 164.77, 167.32, 169.21. MS (EI) m/z 436.42 (M +1).

2-(4-(4-(4-(3,4-dichlorophenyl)thiazol-2-ylamino)-6-(methylamino)-1,3,5-triazin-2-yl)piperazin-1-yl)ethanol (**204D**): Physical state: yellow sticky solid; % yield: 65; m.p.: 218ºC; R_f (silica gel G): 0.51 (ethyl acetate:hexane, 1:1); Solubility: DMSO, ethanol, methanol, chloroform. UVmax (nm): 244.0; FTIR (cm^{-1}) 3388.47 (O-H, str.); 3266.41 (N-H sec, Str.); 2988.98 (C-H Str.); 1238.79, 1187.04 (CN, Ar.) ^{1}H NMR (MeOH): δ ppm: 2.5 (s, 3H, methyl), 2.8 (d, 4H, piperazine), 3.7 (d, 4H, piperazine), 6.4 (s, 1H, Thiazole), 7.5 (d, 2H, phenyl), 7.7 (d, 2H, phenyl). ^{13}CNMR (MeOD): δ,ppm: 40.19, 52.77,125.54, 129.22, 131.90, 132.32, 152.11, 155.44, 165.44, 169.43, 171.32. MS (EI) m/z 479.82 (M +1).

6-(4-methylpiperazin-1-yl)-N2-(4-phenylthiazol-2-yl)-1,3,5-triazin e-2,4-diamine (**7A**): Physical state: brown sticky solid. Yield 62%, m.p.:40^{o}C, R_f value TLC (Hexane:Ethylacetate, 1:1) 0.83, Solubility: Methanol, DMSO, ethanol, chloroform, UV_{max} (nm): 309.0; FTIR(cm^{-1}) 3449.34, 3359.67 (N-H primary, Str.); 1640.02, 851.83(N-H primary, Bend.); 1451.92 (N-H secondary, Bend.); 2949.05 2809.33 (C-H Str.); 1279.57, 1185.64 (CN, Ar.), ^{1}H NMR (MeOD):δ ppm: 2.302 (m, 3H, methylene), 2.41 (d, 4H, piperazine), 3.67 (d, 4H, piperazine), 6.66 (s, 1H, Thiazole), 7.25 (m, 1H, phenyl) 7.35 (d, 2H, phenyl), 7.47 (d, 2H, phenyl). ^{13}CNMR (MeOD): δ ppm: 42.99, 45.15-46.31, 55.21, 6.85, 76.87-77.72, 101.62, 125.77-128.56, 134.67, 150.73, 168.27, MS (EI) m/z 369.12 (M +1).

2-(4-(4-(methylamino)-6-(4-p-tolylthiazol-2-ylamino)-1,3,5-triazin-2-yl)piperazin-1-yl)ethanol (**144C**): Physical state: yellow sticky; % yield: 79; m.p.: 45ºC; R_f (silica gel G): 0.74 (ethyl acetate:hexane, 1:1); Solubility: DMSO, ethanol, methanol; UV_{max} (nm): 233.0; FTIR (cm^{-1}) 3301.66 (O-H, Str.); 1592.84 (N-H sec, Bend.); 2941.19, 2820.99 (C-H Str.); 1303.26, 1262.86 (CN, Ar.) ^{1}H NMR (MeOD): δ ppm: 2.2 (s, 3H, methyl), 2.4 (s, 3H, methyl), 2.7 (d, 4H, piperazine), 3.9 (d, 4H, piperazine), 6.6 (s, 1H, Thiazole), 7.4 (d, 2H, phenyl), 7.6 (d, 2H, phenyl). ^{13}CNMR (MeOD): δ,ppm: 40.22, 52.09, 121.22, 123.22, 125.01, 126.59, 127.71, 147.67, 150.98, 163.02, 166.65, 167.88. MS (EI) m/z 426.21 (M +1).

N2,N2-dimethyl-6-(piperazin-1-yl)-N4-(4-p-tolylthiazol-2-yl)-1,3,5-triazine-2,4-diamine (**171C**): Physical state: yellow sticky; % yield: 73; m.p.: 68ºC; R_f (silica gel G): 0.65 (ethyl acetate:hexane, 1:1); Solubility: DMSO, ethanol, methanol; UV_{max} (nm): 277.0; FTIR (cm^{-1}) 3361.61 (N-H sec, Str.); 1560.21 (N-H sec, Bend.); 2922.17, 2865.72 (C-H Str.); 1367.96, 1230.58 (CN, Ar.) ^{1}H NMR (MeOD): δ ppm: 2.2 (s, 3H, methyl), 2.5 (s, 3H, methyl), 3.2 (d, 4H, piperazine), 3.7 (d, 4H, piperazine), 6.7 (s, 1H, Thiazole), 7.6 (d, 2H, phenyl), 7.8 (d, 2H, phenyl). ^{13}CNMR (MeOD): δ,ppm: 40.29, 51.89, 124.77, 128.33, 129.45,130.98, 150.62, 164.79, 167.19, 168.21. MS (EI) m/z 396.37 (M +1).

3.3 Antimalarial activity evaluation

Table 4: *In -vitro* antimalarial activity of the synthesized molecules

Molecule	% parasitemia (3D7)		% parasitemia (Dd2/RKL2)		IC_{50} (Dd2/RKL2) (µM)	pIC_{50}	Predicted Activity – 2D QSAR	Predicted Activity – 3D QSAR
	5 µg/ml	50 µg/ml	5 µg/ml	50 µg/ml				
201D	98.5	100	71.0	100	3.02	5.52	4.92	5.14
204D	96.0	98.5	69.0	100	2.17	5.66	4.95	5.23
7A	21.5	46.5	17.5*	32.5*	43.8*	4.35	3.94	4.75
144C	31	47	0*	16.5*	38.43*	4.41	4.17	4.03
171C	19	47	0*	13.5*	39.24*	4.40	3.89	4.14

*The strain used was RKL2 (One of the resistant strains of *P. falciparum*)

Molecules that were designed based on the observations from 2D & 3D QSAR had shown good anti-malarial activity. All the molecules were tested against sensitive and resistant strains of *P. falciparum*. Among five molecules 201D and 204D had shown 100% inhibition on resistant strain (Dd2) of *P. falciparum*. Both the molecules were highly active on same resistant strain (Dd2) with IC_{50} of 3.02 and 2.17 µM. As predicted remaining three molecules (7A, 144C and 171C) were depreciated in biological activity when compared to active molecules. *In-vitro* activities of all the synthesized molecules were shown in table 4.

4. Conclusion

The 2D and 3D QSAR models generated by the machine learning approach to understand the relevance of different substitutions on the thiazolyl-triazine core has suggested the presence of favourable secondary amino nitrogen and replacement of the non-favourable hydrophobic atoms with favourable hydrogen bond donors like substituted piperazine on triazine ring. Also, the favourable hydrophobic interaction of the phenyl ring was obtained by substituting the hydrogen with chlorine atoms at third and fourth positions. These findings of the ML approach was further validated by the synthesis of five

compounds, two having chlorine substitution on the phenyl ring whereas the other three was having hydrogen or methyl groups attached to the phenyl ring. The two thiazolyl-traizine derivatives, 201D and 204D proved to be active against resistant strains of *Plasmodium falciparum* and can be further developed as next level anti-malarials to diminish the global burden using the model detailed in this study. This interdisciplinary process of combining medicinal chemistry, supervised machine learning and QSAR had steered the advancement of alternate approaches in antimalarial drug discovery process.

Acknowledgements

The financial support from the Department of Biotechnology (DBT), New Delhi, India is gratefully acknowledged [grant no. BCIL/NER-BPMC/2013-541 (BT/330/NE/TBP/2012 dated 29 April 2013)]. The authors also are thankful to S.A.I.F., Punjab University, Chandigarh, India for providing spectroscopic data; RMRC-Lahoal, Dibrugarh for providing the facilities for carrying out the antimalarial assays; Dibrugarh University for providing the infrastructure and also to Schrodinger for providing the software.

References

[1] Rieckmann K.H., Sax L.J., Campbell G.H., Mrema J.E., 1978. Drug sensitivity of *P. falciparum*: an *in vitro* microtechnique. Lancet. 1, 22–23.

[2] Genther C.S., Smith C.C., 1977. Antifolate studies. Activities of 40 potential antimalarial compounds against sensitive and chlorguanide triazine resistant strains of folate-requiring bacteria and *Escherichia coli*. J. Med. Chem. 20, 237–243.

[3] Gahtori P., Singh A., Ghosh S.K., Das A., Archana U., 2011. Synthesis of some substituted phenylthiazolyl 1, 3, 5-triazine derivatives. Asian J. Chem. 23, 1189–1192.

[4] Gahtori P., Ghosh S.K., Singh B., Singh U.P., Bhat H.R., Uppal A., 2011. Synthesis, SAR and antibacterial activity of hybrid chloro, dichloro-phenylthiazolyl-*s*-triazines. Saudi Pharm. Journ. 20, 35–43.

[5] Gahtori P., Ghosh S.K., 2011. Design, synthesis and SAR exploration of hybrid 4-chlorophenylthiazolyl-***s***-triazine as potential antimicrobial agents. J. Enz. Inhib. and Med. Chem. 1–13.

[6] Gahtori P., Ghosh S.K., Parida P., Prakash A., Gogoi K., Bhat H.R., Singh U. P., 2012. Antimalarial evaluation and docking studies of hybrid phenylthiazolyl-1,3,5-triazine derivatives: A novel and potential antifolate lead for Pf-DHFR-TS inhibition. Ex. Para. 130, 292–299.

[7] Kalita J.M., Ghosh S.K., Sahu S., Dutta M., 2017. Rational design and microwave assisted synthesis of some novel phenyl thiazolyl clubbed s-triazine derivatives as antimalarial antifolate. Fut. J. Pharm. Scs. 3, 11-17.

[8] Sahu S., Ghosh S.K., Gahtori P., Singh U.P., Bhattacharyya D.R., Bhat H.R., 2019. In silico ADMET study, docking, synthesis and antimalarial evaluation of thiazole-1,3,5-triazine derivatives as Pf-DHFR inhibitor. Pharmacolog. Report. 71, 762–767.

[9] Sahu S., Ghosh S.K., Ghoshal A., Kalita J., Gahtori P., Bhattacharyya D.R., 2016. Microwave assisted synthesis, antimalarial screening and structure–activity-relationship exploration of some phenylthiazolyl-triazine derivatives against dihydrofolate reductase. Med. Chem. Res. 25, 2916-2923.

[10] Gahtori P., Pandey R., Kumar V., Ghosh S.K., Das A., Kalita J.M., Sahu S., Prakash A., Bhattacharyya D.R., 2016. Toward resistance-compromised DHFR inhibitors part 1: combined structure/ ligand-based virtual screenings and ADME-Tox profiling. J. Chemomet. 30, 462-481.

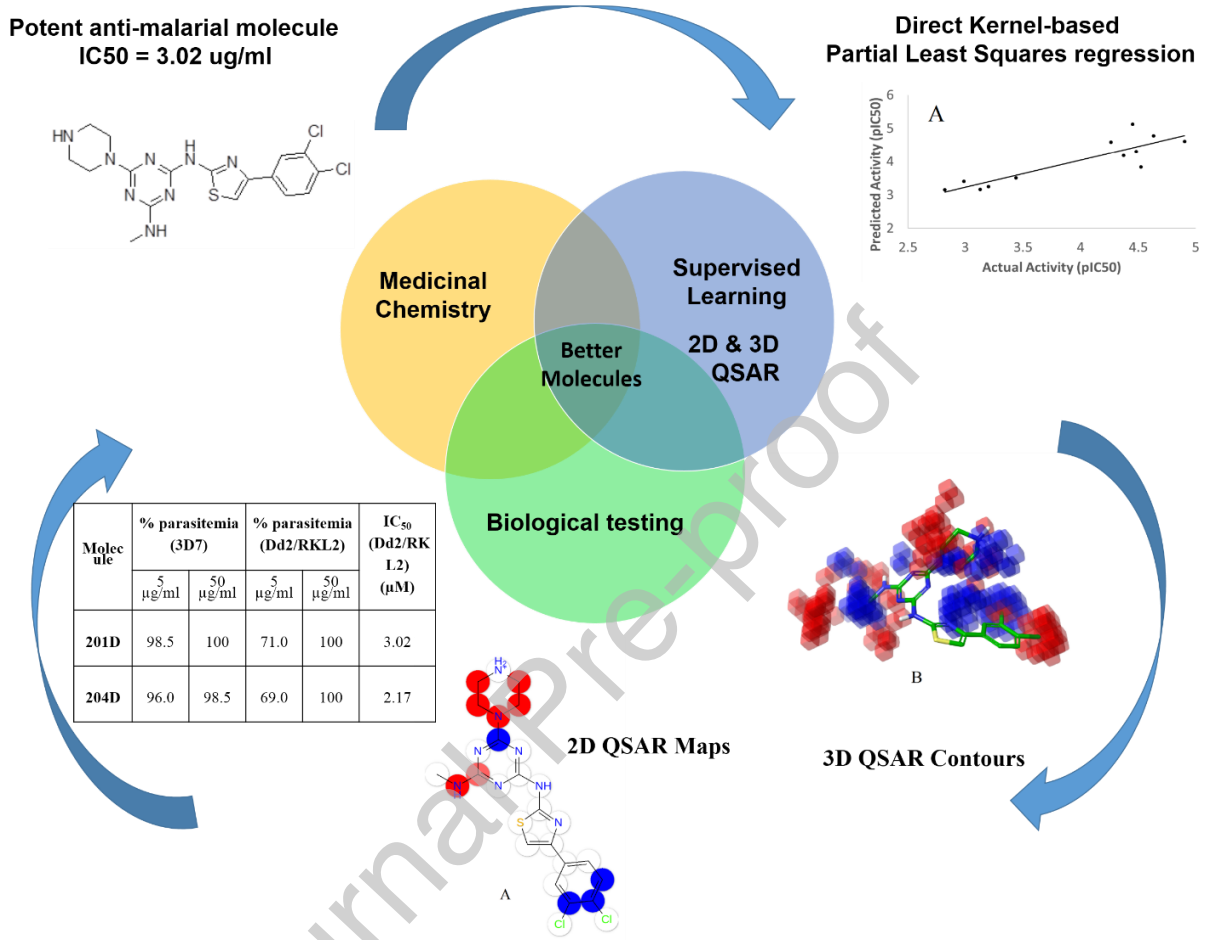
[11] Bender A., Mussa H.Y., Glen R.C., Reiling S., 2004. Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): evaluation of performance. J. Chem. Inf. Comput. 44, 1708–1718.

[12] An Y., Woody S., Steven D.L., 2013. Kernel-Based Partial Least Squares: Application to Fingerprint-Based QSAR with Model Visualization**.** J. Chem. Inf. and Mod. 53, 2312-2321.

[13] Canvas, version 1.5; Schrodinger L.L.C.: New York, 2012.

[14] Raj K.K., Manohar S., Talluri V.R., Rawat D.S., 2015. Insights into activity enhancement of 4-aminoquinoline-based hybrids using atom-based and field-based QSAR studies. Med. Chem. Res. 24, 1136-1154.

Graphical Abstract



Molecule	% parasitemia (3D7)		% parasitemia (Dd2/RKL2)		IC_{50} (Dd2/RKL2) (µM)
	5 µg/ml	50 µg/ml	5 µg/ml	50 µg/ml	
201D	98.5	100	71.0	100	3.02
204D	96.0	98.5	69.0	100	2.17